National Computational Infrastructure for Lattice Gauge Theory

A proposal in response to Office of Science Notice DE-FG02-06ER06-04 and Announcement Lab 06-04: Scientific Discovery through Advanced Computing.

Lead Institution: University of California, Santa Barbara
Santa Barbara, CA 93106

Lead Principal Investigator and DOE Contact: Robert Sugar
Address: Department of Physics
University of California
Santa Barbara, CA 93106
Email: sugar@physics.ucsb.edu
Phone: 805-893-3469

Office of Science Programs Addressed: High Energy Physics and Nuclear Physics

Office of Science Program Office Technical Contacts: Craig Tull and Sidney Coon

Participating Institutions and Principal Investigators:

Physics:

Boston University*, Richard Brower † ‡ and Claudio Rebbi †
Brookhaven National Laboratory*, Michael Creutz † ‡
Columbia University*, Norman Christ † ‡
Fermi National Accelerator Laboratory*, Paul Mackenzie † ‡
Indiana University*, Steven Gottlieb ‡
Massachusetts Institute of Technology*, John Negele † ‡
Thomas Jefferson National Accelerator Facility*, David Richards † and William (Chip) Watson ‡
University of Arizona*, Doug Toussaint ‡
University of California, Santa Barbara*, Robert Sugar † ‡
University of Utah*, Carleton DeTar ‡
University of Washington, Stephen Sharpe †

Computer Science:

DePaul University*, Massimo DiPierro ‡
Illinois Institute of Technology*, Xian-He Sun ‡
University of North Carolina*, Daniel Reed ‡
Vanderbilt University*, Theodore Bapty ‡

* Institution submitting an application
† Project Principal Investigator, Member of Lattice QCD Executive Committee
‡ Institution Principal Investigator

1 Executive Summary

Our long range objective is to construct the computational infrastructure needed for the study of quantum chromodynamics (QCD). Nearly all theoretical physicists in the United States involved in the numerical study of QCD are participating in this effort [1], as are Brookhaven National Laboratory (BNL), Fermi National Accelerator Laboratory (FNAL) and Thomas Jefferson National Accelerator Facility (JLab), and computer scientists at DePaul University, the Illinois Institute of Technology, the University of North Carolina and Vanderbilt University. A very successful start was made under the first phase of the Department of Energy's Scientific Discovery through Advanced Computing Program (SciDAC-1). We propose to build on this success to address new challenges that must be met in order to capitalize fully on the exciting opportunities now available for advancing the study of QCD.

QCD is the component of the Standard Model of elementary particle physics that describes the strong interactions. The Standard Model has been enormously successful; however, our knowledge of it is incomplete because it has proven extremely difficult to extract many of the most important predictions of QCD, those that depend on the strong coupling regime of the theory. To do so from first principles and with controlled systematic errors requires large scale numerical simulations within the framework of lattice gauge theory. Such simulations are needed to address problems that are at the heart of the DOE's large experimental programs in high energy and nuclear physics. Our immediate objectives are to 1) calculate weak interaction matrix elements to the accuracy needed to make precise tests of the Standard Model; 2) determine the properties of strongly interacting matter under extreme conditions such as those that existed in the very early development of the universe, and are created today in relativistic heavy ion collisions; and 3) calculate the masses of strongly interacting particles and obtain a quantitative understanding of their internal structure. The infrastructure we propose to build is essential to achieve these objectives.

The bulk of our effort in SciDAC-1 was devoted to software development, and that will continue to be the case under this SciDAC-2 proposal. Under SciDAC-1 a QCD Applications Programming Interface (QCD API) was developed, which enables lattice gauge theorists to make effective use of a wide variety of massively parallel computers, including those with switched and mesh architectures. The QCD API was optimized for the custom designed QCD on a Chip (QCDOC) computer, and for commodity clusters based on Pentium 4 processors. Under this proposal, optimized versions of the QCD API will be created for clusters based on multi-core processors and Infiniband communications networks, and for the Cray XT3, the IBM BlueGene/L and their successors. The QCD API will be used to enhance the performance of the major QCD community codes and to create new applications. A QCD physics toolbox will be constructed which will contain sharable software building blocks for inclusion in application codes, performance analysis and visualization tools, and software for automation of physics work flow. New software tools will be created for managing the large data sets generated in lattice QCD simulations, and for sharing them through the International Lattice Data Grid consortium. A common computing environment will be developed for the dedicated lattice QCD computers at BNL, FNAL, and JLab. Work on multi-scale algorithms recently begun in collaboration with members of the Terascale Optimal PDE Simulations (TOPS) Center will be extended.

The lattice QCD infrastructure effort has included the development of hardware as well as software, because for the study of QCD it has proven more cost effective to build specialized computers than to make use of general purpose supercomputers. We have pursued both customized clusters constructed from commodity components and the development of fully customized computers. Research and development work on commodity clusters was carried out under our SciDAC-1 grant at FNAL and JLab. The experience gained with these prototype clusters will enable us to build highly cost effective terascale clusters in the coming year. In parallel with SciDAC-1, but funded separately by the DOE, a 12,288 processor QCDOC computer was constructed at BNL for use by the U.S. lattice QCD community. In SciDAC-2 we propose to continue to track the evolving commodity and semi-commodity marketplace and to undertake design of a fully customized successor to the QCDOC. A four year Lattice QCD Computing Project began on October 1, 2005 with funding from the DOE's High Energy Physics and Nuclear Physics Programs. The purpose of this Project is to construct and operate dedicated computers for the study of QCD. Both the hardware research and development and the software development we propose are critical to the success of this Project and to research in lattice QCD in the U.S.

2 Physics Goals

2.1 Tests of the Standard Model

Despite its extraordinary success, the Standard Model is believed to be only the low energy (long distance) limit of a more fundamental theory. Therefore, a major component of the experimental program in high energy physics is devoted to making precise tests of the Standard Model in order to determine its range of validity and search for indications of new physics beyond it. Many of these tests require both accurate experiments and accurate lattice QCD calculations of the effects of the strong interactions on weak interaction processes. In almost all cases, the precision of the tests are limited by the uncertainties in the lattice calculations, rather than in the experiments. Our objective is to bring the lattice errors down to, or below, the experimental ones.

The greatest challenge to performing accurate numerical calculations of QCD is to include the full effects of vacuum polarization due to light (up, down and strange) quarks. Significant progress has been made in meeting this challenge during the past five years through the use of improved formulations of QCD on the lattice and through rapid growth in the computing resources available to the field [2, 3, 4]. Among the notable results have been calculations of the leptonic decay constants of the π and K mesons [5] and mass splittings in the charmonium [6] and bottomonium [7] spectra to an accuracy of 3% or better; the first determination of the light quark masses to fully include their vacuum polarization effects [8, 9]; the calculation of the strong coupling constant [10] and the Cabibbo-Kobayashi-Maskawa (CKM) matrix element V_{us} [11, 5] to the same accuracy as their experimental determinations. The lattice gauge theory community has moved from the validation of techniques through the calculation of quantities that are well known experimentally to the successful prediction of quantities that had not previously been measured. Three cases in which predictions were subsequently confirmed by experiment were the calculations of the leptonic decay constant [12] and semi-leptonic form factors [13] of the D meson, and the mass of the B_c meson [14]. The decay constants and form factors for B mesons play important roles in tests of the Standard Model, but are very difficult to measure experimentally. The lattice calculations are similar for D and B mesons, since only the masses of the heavy quarks change. Thus, the successful calculations for D mesons provide important validation of those for B mesons which are now in progress.

Measurement	CKM Matrix Element	Hadronic Matrix Element	Non-Lattice Errors	Lattice Errors 2004	Lattice Errors Current	Lattice Errors 6.0 TF-Yr	Lattice Errors 40. TF-Yr
ε_K ($\bar{K}K$ mixing)	$\mathrm{Im}\, V_{td}^2$	$\hat{B}_K$	9%	20%	12–20%	5%-8%	3%-4%
ΔM_d ($\bar{B}B$ mixing)	$\lvert V_{td}\rvert^2$	$f_{B_d}^2 B_{B_d}$	6%	30%	22%	8%–10%	6%–8%
$\Delta M_d/\Delta M_s$	$\lvert V_{td}/V_{ts}\rvert^2$	ξ^2	—	12%	8%	6%	3%–4%
$B \to \binom{\rho}{\pi} l\nu$	V_{ub}	$\left\langle \binom{\rho}{\pi} \right\rvert (V-A)_\mu \lvert B\rangle$	7%	15%	14%	5.5%–6.5%	4%–5%
$B \to \binom{D^*}{D} l\nu$	V_{cb}	$\mathcal{F}_{B\to\binom{D^*}{D}l\nu}$	2%	4.4%	3%	1.8%–2%	1%–1.4%

Table 1: The impact of improved lattice QCD calculations on the determination of CKM matrix elements.

The results quoted above indicate that we are in a position to make very significant progress over the next five years. The current lattice and experimental uncertainties in some key quantities are shown in Table 1, along with the reduction in lattice errors expected as more computational resources become available, as well as expected improvements in ancillary theoretical calculations of operator normalization factors. All quantities in the table have had first calculations which fully include the effects of vacuum polarization due to light quarks. The error estimates in Table 1 are based on our experience with the improved staggered formulation of lattice quarks, as were the successful calculations cited above, with the exception of the ε_K estimates which are based on domain wall quarks as well.

2.2 Matter under extreme conditions

At very high temperatures and/or densities, one expects to observe a phase transition or crossover from ordinary strongly interacting matter to a plasma of quarks and gluons. A primary motivation for the construction of the Relativistic Heavy Ion Collider (RHIC) at BNL was to observe the quark–gluon plasma and determine its properties. During the early development of the Universe matter was in the plasma state, and the quark-gluon plasma may be a central component of neutron stars today. The behavior of strongly interacting matter in the vicinity of the phase transition or crossover is inherently a strong coupling problem, which can only be studied from first principles through lattice gauge theory calculations. Among the issues that can uniquely be addressed by lattice calculations are the nature of the transition, the temperature at which it occurs, the properties of the plasma, and the equation of state. Indeed, it is the lattice that has given us the best estimates of the temperature of the deconfinement transition [15]. Lattice results will continue to be crucial to the interpretation of ongoing heavy–ion experiments in the United States and Europe.

A major goal of our research program is to investigate the properties of matter under the extreme conditions of high temperature and high density. Important progress has been made in the last several years [15] in determining the transition temperature [16, 17], the phase diagram [17] and the equation of state [18, 19, 20, 21]. However, as the deconfinement process occurs at a temperature of order 175 MeV, a new scale is introduced, as well as new potential lattice artifacts. Thus, these calculations are computationally quite demanding. Those at zero baryon density are well understood theoretically, and only require sufficient computational resources to reach high precision results; but, calculations at non-zero baryon density are at a much earlier stage of development.

The finite density problem introduces algorithmic issues that remain unresolved and require exploratory work on new ideas such as calculations at fixed quark number rather than fixed chemical potential. Moreover, several new approaches to this long-standing problem [18, 19, 20] have been suggested that are applicable at high temperature and small values of the baryon density, the regime relevant to the RHIC experiments. Also, at vanishing baryon density, new exact algorithms have been developed for staggered fermions. These new techniques have to be explored and implemented into existing code packages.

The main goal of the ongoing studies at zero baryon density is to extend calculations of the transition temperature, the equation of state and the properties of the high temperature phase to an almost realistic quark mass spectrum on large lattices with small lattice spacings. This will allow a controlled extrapolation to the continuum and thermodynamic limits. This research effort is using a large portion of the resources provided by the DOE QCDOC supercomputer at Brookhaven. Specifically, the lattice group at BNL and the MILC collaboration use improved staggered fermion actions (p4-action, asqtad) with smeared links to reduce flavor symmetry breaking and the cut-off distortion of thermodynamic observables [22].

Code packages for studies of thermodynamics at non-zero baryon density are implemented in these calculations. The specific approach used by the BNL and MILC groups is based on a high order Taylor series expansion. This will allow exploration of the QCD phase diagram also at high temperature and non-zero baryon density.

A somewhat different computational set-up is required to study hadron properties at high temperature. The goals here are two-fold. In the heavy quark sector one wants to understand the stability of charmonium states in the finite temperature plasma and determine the temperature at which these states get dissolved [23]. In the light quark sector, the emphasis is on calculating thermal dilepton and photon rates. Currently these calculations are being performed on large quenched lattices using Wilson fermions. These calculations will be extended to improved Wilson fermions, which reduce the distortion effects resulting from so-called Wilson doublers. Moreover, preparations for the first exploratory studies of these effects with dynamical quarks are underway.

2.3 Structure and interactions of hadrons

A major scientific goal of our collaboration is to achieve a quantitative, predictive understanding of the structure and interactions of strongly interacting particles (hadrons) from lattice QCD. This will achieve key objectives of the DOE Strategic Plan and the Nuclear Science Long Range Plan, which respectively highlight the goals of developing a quantitative understanding of how quarks and gluons provide the binding and spin

of the nucleon based on QCD and of connecting the observed properties of nucleons with the underlying theoretical framework provided by QCD. Hadronic observables calculated from first principles are directly relevant to experiments at Bates, JLab, RHIC-spin, SLAC,and FNAL, and will have significant impact on future experiments at the JLab 12 GeV upgrade and electron-ion collider.

Past accomplishments have established the methodology and laid the groundwork for hadron structure and spectroscopy calculations. Since the cost of full QCD calculations in a volume large enough to contain a pion grows roughly as $m_\pi^{-7} - m_\pi^{-9}$, initial calculations were restricted to the "heavy pion" domain of pion masses in excess of 500 MeV. In this domain, form factors, the lowest three moments of quark, spin, and transversity distributions, and generalized form factors corresponding to the lowest three moments of generalized parton distributions have been calculated [24, 25]. Salient achievements include separating the contributions of the quark spin and orbital angular momentum to the nucleon spin [24] and observing strong dependence of the transverse size on the nucleon on the longitudinal momentum fraction [25]. The transition form factor between the nucleon and Delta has been calculated to explore the role of deformation [26, 27]. In spectroscopy, techniques to calculate extended sources within the appropriate representation of the hypercubic group have been developed and utilized to calculate ground and excited states in each symmetry channel [28, 29] and pentaquark states were calculated using a complete set of local sources [30].

An essential step toward the chiral regime with light quarks has recently been taken using a hybrid calculation combining computationally economical staggered sea quark configurations generated by the MILC collaboration and domain wall valence quarks that have lattice chiral symmetry. The axial charge has recently been calculated for pion masses as light as 350MeV. Since this is in the regime of applicability of chiral perturbation theory, analytic expressions for the mass and volume dependence were used to extrapolate to the physical pion mass and infinite volume, obtaining the axial charge to a precision of 6.8% and in agreement with experiment. Hybrid calculations of the pion form factor were also performed [32]. In a first step in studying hadron-hadron interactions, the $I = 2$ $\pi - \pi$ scattering length [33] and nucleon-nucleon ${}^1S_0$ and ${}^3S_1 - {}^3D_1$ scattering lengths [34] have also been calculated in this chiral regime.

Building on this solid foundation, we propose an extensive program of precision calculations of the hadron observables described above and exploratory calculations of more demanding observables. Using MILC configurations at lattice spacings of 0.12, 0.09 and 0.06 fm and pion masses down to 250 MeV, form factors, moments of quark, spin, and transversity distributions, generalized form factors, and transition form factors will be calculated with careful control of the errors associated with the lattice spacing, lattice volume, and quark mass. When computational resources permit, the hybrid combination of staggered and chiral quarks will be replaced by fully consistent sea and valence quarks.

Important new algorithms and observables will also be explored. The spectroscopy of so-called "missing" baryon resonances and of mesons with exotic quantum numbers has great potential impact on our understanding of QCD, and on the current and future spectroscopy program at JLab. Exploration of the baryon and meson spectrum into the chiral regime requires further development. Multi-hadron operators and the use of a range of lattice volumes will be used to study the properties of unstable resonances. Stochastic all-to-all quark propagators with dilution and exactly-determined low eigenvectors will be used to facilitate both these spectroscopy calculations and the calculation of disconnected diagrams for hadron structure observables. Nucleon-nucleon scattering lengths will be calculated in the chiral regime and the static potentials between heavy-light hadrons will be explored. Since past efforts to calculate important gluon observables in hadrons have been overwhelmed by the large fluctuations of the gluon field, new approaches will be investigated. Hadron calculations in the chiral regime also open the door to understanding the physical origin of observed structure, and the role of mechanisms such as diquark correlations and the quark zero modes associated with topological excitations will be explored. This combination of well-understood observables that can be calculated with confidence given the requisite resources and more speculative exploration of new physics offers exciting opportunities for the fundamental understanding of hadronic physics.

2.4 Lattice quarks and other physics directions

In recent years, much of the progress in the numerical study of QCD has come about through the use of improved formulations of quarks on the lattice. There are a number of different formulations that appear to be promising, each of which has its advantages and disadvantages. Most of the work cited above made use of the staggered formulation of lattice quarks. This formulation has the advantage of enabling simulations

at quite small quark masses with current computers, but in order to have the correct number of quarks in the continuum limit one must perform simulations with the fourth-root of the quark determinant. It has been suggested that taking the fourth-root of the determinant at finite lattice spacing might give rise to unphysical non-localities that persist in the continuum limit [35]. The excellent agreement of existing results with experiment, as well as a growing body of direct discussions of the issue [36, 37], give us confidence that no fundamental problem exists. However, further work is warranted, and in progress. We have started an extensive set of simulations with domain wall quarks, which is likely to continue through most, or all, of the proposed grant [4]. This formulation has the advantage of having nearly exact chiral symmetry on the lattice, but requires significantly more computing resources than staggered quarks for the same parameter regime. Studies with domain wall quarks will increase the range of quantities that can be computed, will provide critical tests of the staggered quark results, and in the long run may increase the accuracy of those results [38].

The methods used for QCD can also be adapted for other strongly coupled theories. Noteworthy examples include QCD with a large number of colors, where it may be possible to build a bridge to analytic methods based on string theory and the AdS/CFT correspondence; a strongly coupled Higgs sector; and proposed models for physics beyond the standard model involving strongly coupled gauge interactions such as supersymmetry and "little Higgs" models. These other applications are generally more challenging than QCD, and work is at an early stage. We expect, however, that an increasing fraction of the US community will work on such theories during the next five years.

3 SciDAC-1 Software: The QCD Applications Programming Interface

Under its SciDAC-1 grant, the U.S. lattice gauge theory community has created a unified program environment that enables its members to achieve high efficiency on terascale computers. Among the design goals were to enable users to quickly adapt codes to new architectures, easily develop new applications and incorporate new algorithms, and preserve their large investment in existing codes. These goals were achieved through the development of the QCD Applications Programming Interface (QCD API), which is illustrated in Fig. 1.

All of the fundamental components of the QCD API have been implemented and are in use on the U.S. QCDOC hardware at BNL, on both the switched and mesh architecture Pentium 4 clusters at FNAL and JLab, and on a number of general purpose supercomputers. The QCD API is being used by a growing number of physicists in the U.S. and abroad. The software code and documentation can be found at the USQCD http://www.usqcd.org/usqcd-software. Here we briefly describe each of its components.

The QCD API has a layered structure which is implemented in a set of independent libraries. Level 1 provides the code that controls communications and the core single processor computations. To obtain high efficiency on terascale facilities, much of this layer may have to be written in hardware specific assembly language. However versions exist in C and C++ using MPI for transparent portability of all application codes.

Message Passing: QMP defines a uniform subset of MPI-like functions with extensions that (1) partition the QCD space–time lattice and map it onto the geometry of the hardware network, providing a convenient abstraction for the Level 2 data parallel API (QDP); (2) contain specialized routines designed to access the full hardware capabilities of the QCDOC network and to aid optimization of low level protocols on networks in use and under development on clusters. There is a basic test suite to verify each implementation.

Linear Algebra: All lattice QCD calculations make use of a set of linear algebra operations in which the basic elements are three–dimensional complex matrices, elements of the group SU(3). These operations are local to lattice sites or links and do not involve inter–processor communications. We have collected them into a single Level 1 library called QLA. The QLA routines can be used in combination with QMP to develop complex data parallel operations in QDP or in existing C or C++ code. The C implementation has about 19,000 functions generated in Perl, with a full suite of test scripts. The C++ implementation makes considerable use of templates, and so contains only a few dozen templated classes (the required specific classes are generated on demand by the compiler). For both C and C++ it is important to optimize the code for the most heavily used linear algebra modules.

Data Parallel Interface: Level 2 (QDP) contains data parallel operations that are built on QMP and QLA. The C implementation is being used to improve performance of the MILC code, a large, publicly available suite of applications. Despite the fact that the MILC code has been carefully optimized over its fifteen year lifetime, rewriting computationally intensive subroutines in QDP makes a significant improvement in its performance. Chroma, an entirely new application code base, has been written *di novo* in the C++ implementation of QDP. QDP allows extensive overlapping of communication and computation in a single line of code. By making use of the QMP and QLA layers, the details of communications buffers, synchronization barriers, vectorization over multiple sites on each node, etc. are hidden from the user.

Level 3 Subroutines: A very large fraction of the resources in any lattice QCD simulation go into a few computationally intensive subroutines, most notably the repeated inversion of the Dirac operator, a large sparse matrix. To obtain the level of efficiency at which we aim, it is necessary to optimize these subroutines for each architecture. For example, on the QCDOC, the assembly coded inverter for the Domain Wall and Asqtad quark actions, the two quark formulations that are being used in initial work, is as high as 42% and 45% of peak, respectively. (The precise performance depends on the number of lattice sites assigned to each processor). These percentages correspond to total sustained performances of 4.1 and 4.4 teraflop/s for the full 12,288 processor machine. Level 3 codes written with SSE2 instructions achieve up to 3.0 gigaflop/s per processor for the most recent cluster built at JLab, which has 3.0 GHz dual core Pentium 4 processors.

Data Management: A very large fraction of the computing resources used in lattice QCD calculations go into Monte Carlo simulations that generate representative configurations of the QCD ground state. The same configurations can be used to calculate a wide variety of physical quantities. Because of the large resources needed to generate configurations, the U.S. lattice community has agreed to share all of those that are generated with DOE resources. To enable this sharing we have created standards for file formats, and written an I/O library (QIO) that adheres to them. We are charter members of the International Lattice Data Grid (ILDG), which is setting a basic set of meta-data and middleware standards to enable international sharing of data. By June 2006, the U.S. lattice gauge theory community will be fully capable of archiving and retrieving data on the ILDG.

4 SciDAC-2 Software

The full benefits of the SciDAC-1 infrastructure are just beginning. To capitalize on the accomplishments to date will require continued work on porting, optimization, testing and distribution of software libraries. In addition, there is a new set of requirements and challenges as we prepare for the petaflop/s era. We propose to extend the QCD API and its related libraries under SciDAC-2 to meet these challenges. As a guide to the discussion below, we summarize the proposed API in Fig. 1.

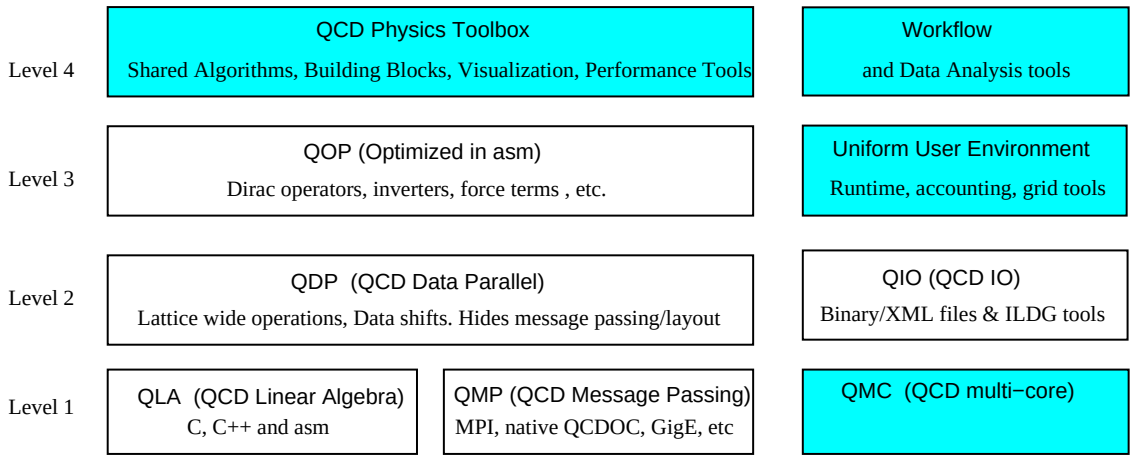

Figure 1: Proposed SciDAC-2 QCD API — The SciDAC-1 components are shown in white, and the new SciDAC-2 components in aqua.

4.1 Machine specific software

The basic MPI and C/C++ code is highly portable, but to achieve high performance may require machine specific software for both the Level 1 and 3 routines. Such software has been written under SciDAC-1 for the QCDOC and clusters based on single core Pentium 4 processors. Machines that will be targeted in the first stages of this grant are clusters based on multi-core processors and Infiniband communications fabrics, and the Cray XT3, the BlueGene/L and their successors. Attention will also be paid to emerging technology, such as the Cell processor, so that we are ready to take advantage of any major new developments that might emerge.

QMC: Threaded libraries for multi-core processors: All of the principal manufacturers of commodity microprocessors, including Intel, AMD, and IBM, have started the move toward multi-core processors in the last two years. The latest SciDAC-1 prototype cluster is based on an Intel dual core microprocessor, and the first large scale cluster constructed under the LQCD Computing Project will make use of a multi-core processor as well. By 2007, the majority of processors sold to the commodity market are expected to be dual core, with a planned movement to quad and higher cores. The BlueGene/L has dual core PowerPC processors, and we anticipate a rapid expansion in the number of cores in the BlueGene/P and the QCDOC-2 described below.

In the short run, lattice QCD application codes can take advantage of multi-core microprocessors by simply treating the cores as independent processors. That is, a lattice QCD application implemented with QMP or MPI can run using a separate process on each of the cores. Message passing between the processes running on the cores relies on the shared memory structures provided by MPI implementations for SMP systems. Although communications between the processes requires copying data from one process to the shared memory and back from the shared memory to the second process, the scaling observed using this technique is very encouraging. However lattice QCD codes tend to be memory bandwidth limited, so we anticipate that threaded code will be necessary for peak performance. Unfortunately, standard implementations of POSIX threads often have high overhead for locks, or in some cases may not be available on specialized architectures. We therefore propose to develop a new light-weight Level 1 multi-core library or threaded interface standard (QMC). This is conceptually on the same level as QLA and QMP. It will provide an abstraction for threads that can be implemented for portability in the POSIX standard, but will have native implementations for highest performance and for unconventional multi-core architectures.

The QDP and Level 3 codes based on QMC will be improved by avoiding the memory copies used by MPI to pass messages between processes. We will begin by carefully studying the options. Several techniques are available to avoid the memory copies: (i) Have independent processes on each of the cores use the same shared memory area to store the sub-lattice, with each process performing calculations on a fraction of the sub-lattice; (ii) Use OpenMP or a similar parallel compiler to implicitly thread key code loops. In this case, one process runs on a given machine, with one thread spawned per core, performs calculations on a fraction of the sub-lattice; (iii) Use explicit multi-threading with a thread on each core handling a fraction of the sub-lattice. These techniques vary in difficulty and in level of effort required. Further, these optimizations may be done at either the API library level (QLA, QDP), or at the application level (Level 3), or both. Modeling and software prototyping will be required to investigate the costs and benefits of the approaches.

QMP: Native implementations for Infiniband and BlueGene: The latest lattice QCD clusters constructed at JLab and FNAL under our SciDAC-1 grant, and the initial ones to be constructed in the LQCD Computing Project are based on Infiniband fabrics. For lattice QCD codes, Infiniband delivers superior price/performance. It offers the highest communications bandwidth between computers available on the market and exhibits low short-message latencies, both critical for lattice QCD applications.

The Infiniband software stack provides several communications protocols, including TCP/IP, channel-based communications (remote direct memory access, or RDMA), and message-based communications (verbs application program interface, or VAPI). The later two, RDMA and VAPI, deliver the best performance in terms of highest bandwidth, lowest latency, and lowest burden to the host processor. Two open source MPI implementations are available which are based on combinations of RDMA and VAPI: MVAPICH, from the Ohio State University, and MPICH-VMI, from the National Center for Supercomputing Applications.

The JLab and FNAL Infiniband clusters currently rely on the implementation of QMP that uses MPI for the underlying communications. Simple benchmarks show that for the message sizes of interest on lattice QCD codes, communications using native RDMA and VAPI calls have lower latencies than those using MVA-

PICH or MPICH-VMI. Careful evaluation of lattice QCD codes using computationally intensive kernels, such as the conjugate gradient inversion routines, will be used to infer whether a "native" QMP over RDMA and/or VAPI has a sufficient performance improvement over an MPI version to warrant the manpower for a full implementation. This software prototyping would be a continuation of work started under the SciDAC-1 grant.

The BlueGene/L, a direct descendant of the QCDOC, has considerable potential for the study of QCD. Given the plan to install a large BlueGene/P at Argonne National Laboratory, it seems worthwhile to develop a native version of QMP for the BlueGene line. We propose to use the single tower BlueGene/L's a Boston University and Massachusetts Institute of Technology, and our close collaboration with IBM in this project.

QLA: Optimized Linear Algebra Routines: As indicated above, it is important to optimize the most heavily used linear algebra routines. This optimization is common to and can be shared between the C and C++ implementations. The approach depends on the specific processor being used.

A limited number of QLA routines have been optimized for the Pentium 4 processors using SSE instructions. The new Intel-based clusters at JLab and FNAL can be run in either 32-bit ("IA32") or 64-bit (x86-64) mode. Preliminary single node testing in the x86-64 mode indicates improved performance in the non-SSE portions of the code. The existing SciDAC-1 QLA library code compiles and runs correctly in the 64-bit environments. The SSE optimizations in QLA can be further improved by taking advantage of the larger register file (16 SSE registers, compared to 8 in the 32-bit mode). We propose to do so.

GNU and commercial compilers now generate SSE code under high optimization levels. Because the compilers are not aware of the registers used by the QLA inline-SSE codes, there can be register conflicts and as a result, incorrect results generated. The QLA SSE codes currently use inline gcc assembler macros. These routines should be rewritten so that register conflicts with compilers no longer occur. Further, the 64-bit environment no longer uses the stack to pass operands and results, but instead uses the larger register files available in 64-bit mode. To simplify the maintenance of the existing QLA SSE codes, the inline assembler macros will gradually be replaced with GNU assembler code.

It is important to optimize key QLA routines for the Opteron processors used in the Cray XT3. Initial experience indicates that the large XT3s at Oak Ridge National Laboratory (ORNL) and the Pittsburgh Supercomputer Center (PSC) are superb tools for the study of QCD, and with its planned upgrade, the ORNL machine has the potential to become the single most powerful computer available to lattice gauge theorists. Although the C version of our codes obtains good performance on the XT3, approximately 800 megaflop/s per processor, the SSE coded routines do not provide the boost in performance seen on Intel processors. Collaboration members are discussing this issue with Cray, and a concerted effort to optimize QLA routines for the Opteron appears warranted. Another reason for investigating Opteron-specific optimizations is that these processors have proven to be very cost competitive, and may become components of clusters built in the LQCD Computing Project. This optimization work is related to the 64-bit work on Intel processors described above.

So far much of the optimized QLA code has been hand written, either directly in assembly code or as the input for an assembly code writing tool such as BAGEL [39]. The rest of the QLA routines are automatically generated in C or C++ code by Perl scripts or expression templates. It would be very beneficial to combine these two steps to facilitate a more rapid optimization of any desired linear algebra module. This requires studying existing techniques for code optimization, such as extending the expression template techniques, and then developing our own tools that will assist in generating optimized code. Initially, we propose to start a feasibility study to automate the generation of QLA by developing a prototype of a generic code generating tool with a back end for the BlueGene/L processor, adding support for more architectures as the technology matures.

QOP: Optimized Level 3 Routines: As previously indicated, the bulk of the floating point operations in any lattice QCD calculation are concentrated in a few routines. Because of the very large computational resources involved, it is worthwhile to hand code these routines for the major platforms that will be used by our field. Under SciDAC-1 the primary focus was on inverters for the Dirac operator on Pentium 4 based clusters and the QCDOC. Under SciDAC-2 the work on Level 3 routines needs to be extended in two directions. First, Level 3 inverters need to be written for the three new platforms which we expect to play major roles in our research: clusters based on multi-core processors and Infiniband communications fabrics, and the Cray XT3, BlueGene/L and their successors. Second, for some improved actions, routines other than the Dirac inverter take enough computing resources to warrant Level 3 coding. The fermion force

routine in the Asqtad is a prime example. A Level 3 routine has recently been written for it on the QCDOC, and this work needs to be extended to other platforms and other routines.

4.2 Infrastructure for physics applications

The development of a common QCD API had to be performed while preserving the large investment in application codes and maintaining a continuous production environment for applications. There are three large scale, freely available application code suites developed by members of the U.S. lattice gauge theory community:

- MILC: The MILC code is an integrated package of some 150,000 lines of scientific application codes and a library of generic supporting codes. It has been in use and freely available to the public since the early 1990's, and is widely used outside the MILC Collaboration. It is written in C, and can be compiled with either the QMP or MPI message passing libraries. It can be downloaded from http://www.physics.utah.edu/~detar/milc/.
- CPS: The Columbia Physics System software begun in 1995 is a comprehensive lattice QCD code primarily used by Columbia, BNL, RIKEN-BNL Research Center and UKQCD lattice theorists. It is written in C++ and targeted for the QCDSP and QCDOC computers. It is also capable of running on clusters, using either QMP or MPI for message passing. It can be downloaded from http://qcdoc.phys.columbia.edu/chulwoo_index.html.
- Chroma: Chroma is a new application code written entirely in C++ on top of the SciDAC-1 QCP API. It is being developed by JLab along with major U.S. and international collaborations. It can be downloaded from http://www.jlab.org/~edwards/chroma/.

Support of the QCD API

Integration and optimization of QCD API: The three major application codes have different designs and application foci. Indeed the basic structure of the QCD API benefited tremendously from the collective experience of the developers of these three different, highly optimized and portable QCD codes. We are in the middle of an evolutionary process of bringing the full benefits of the SciDAC-1 QCD API to them, and propose to accelerate this process under SciDAC-2. This work will include writing additional Level 3 routines callable from all three codes, greater integration of the API into the MILC code, and expansion of Chroma. In addition, we propose to develop a new Physics Toolbox (Level 4), a set of building blocks needed by the entire community to develop new applications and algorithms. Finally, we propose to develop a set of data analysis tools that will enable lattice gauge theorists to handle efficiently the very large data sets they are producing.

Documentation: There is clearly a need for additional documentation of both the QCD API and the publicly available applications codes. As the QCD API moves from the development stage of SciDAC-1, where users were either developers or close colleagues of them, to SciDAC-2 with a rapidly expanding user community, the need for adequate documentation is magnified. Some excellent documentation exists for critical components of the API, but a uniform and complete set is now urgent. Similarly, each of the application codes listed above has a large, highly distributed user community. As these communities grow well beyond the groups that developed the software, the need to upgrade the existing documentation does as well. Three levels of documentation are needed for each software component: documentation for installation, a user's guide for running the software, and a developer's guide for extending the software.

Documentation is a necessary part of releasing quality software. Unfortunately it is also burdensome to write, is difficult to maintain (especially in the case of highly distributed development) and requires substantial expertise (the author of the documentation has to actually know the software components quite well in order to give an accurate description). We propose to undertake a substantial upgrade of the documentation of both the QCD API and the application codes under SciDAC-2.

Testing: We propose to build a comprehensive test framework for all of the QCD API libraries. Testing frameworks in general and test driven development in particular tend to produce cleaner code and reduced coupling between software components. A test environment is needed for nightly builds for codes directly

from the source code repository and should target many different architectures, such as single node workstations, clusters and the QCDOC. The test system should also provide a (nightly) regression test framework to insure correctness. Finally, API tests that verify implementation can determine, among other things, whether calls to optimized Level 3 routines reproduce those to standard C or C++ code.

QCD Physics Toolbox

We propose to construct a QCD physics toolbox which will contain a set of basic software building blocks and tools to aid in the development of application codes, algorithm studies and data analysis. Our objective is to enable users to focus on physics by minimizing the coding effort needed to explore it. We believe that this toolbox has the potential to greatly expedite the development of new application codes and new algorithms.

Shared algorithms and Building blocks: A considerable amount of software can be shared among applications. This includes commonly used routines for reunitarization, gauge fixing, the evaluation of low lying eigenvalues of the Dirac operator, and a host of measurements. It also includes more specialized and intricate routines, such as the determination of the fermion force for improved actions with non-nearest neighbor gauge links, and important new algorithms, such as RHMC. We propose to collect these into the toolbox, from which they can be called by any application code that conforms to the QCD API. A few of the tools that have substantial impact on performance, such as the fermion force in the Asqtad action, should be coded at Level 3, but most can be coded directly in C or C++ on top of the Level 2 QDP/QDP++ interface. We designate this new set of common building blocks and algorithms Level 4 to distinguish them from the relatively few instances in which it is worthwhile to write hand-coded Level 3 routines.

In addition to its role in expediting the development of application code, the Level 4 software will provide important support for rapid exploration and testing of new algorithms. A persistent problem in algorithm research is that to test the efficacy of a new approach often requires simulations and benchmarking on systems of a size used in state of the art of the physics calculations. Thus, the ability to rapidly produce high performance code to test new algorithms is extremely valuable. We therefore propose to build Level 4 routines that will further enhance the ability of the QCD API to support algorithm research.

Graphics and Visualization: Another component of the Level 4 toolbox will be a set of graphics routines that can be called from code that conforms to the QCD API standards. Lattice QCD computations comprise multiple steps, creating very large datasets, but the final result is typically encompassed in a small set of numbers with the analysis performed in an automated way. While an automated procedure may be beneficial in efficiency, the ability to visualize the data being analyzed is important both as an aid to the analysis, and as a means of acquiring insight into the physics. Visualization of lattice data has already provided important insights into QCD: pictures of the four-dimensional action densities and topological charge have revealed the complexities and structure of the QCD vacuum, the energy densities between a heavy quark and anti-quark, and between three heavy quarks, have shown the emergence of flux tubes.

Crucial to the success of the graphics-visualization initiative will be a close collaboration between physicists to devise and interpret visualization of physically important quantities, and computer scientists to provide the appropriate visualization toolbox. Questions that visualization might address are many: can we understand how flux-tube formation observed with infinitely heavy quarks extends to hadrons where one or more of the quarks is light; what is the distribution of charge within a nucleon; can we display the distribution of spin and magnetism within a hadron? In the longer term, can we visualize the interactions of hadrons?

Currently, no general-purpose package is available tailored to the display of lattice data. Thus a software package will be developed with a general GUI capable of reading a set of four-dimensional lattice quantities, and taking their ensemble average; performing a projection into a real four-dimensional vector; interpolating the 4-D vector into a continuous four-dimensional field; taking three-dimensional slices of a four-dimensional field; displaying the data using density plots, iso-surfaces, and 2-D projections; and displaying the evolution of data, both in simulation time for four-dimensional quantities, and as the evolution of three- and two-dimensional slices in the remaining coordinates.

The software will support two types of plug-ins: type-1 plug-ins that perform specific physics measurements and output a real 4-D vector, and type-2 plug-ins that take the interpolated 3-D field and generate specific types of plots.

Most of the research underlying this project will consist of identifying a set of physical measurements suitable to be implemented as type-1 plug-ins. The visualization techniques for the type-2 plug-ins are very

similar to standard techniques used for representation of 3-D geophysical data and, when possible, we will incorporate existing libraries into the development of our plug-ins.

The system will be developed in C++ and take advantage of existing graphics and visualization libraries such the Trolltech QT libraries and the Visualization Tool Kit (VTK) library. The plug-ins will be callable from C or C++ code conforming to the QCD API, and will form another component of our Level 4 QCD Toolbox. The system will be capable of reading datasets in the SciDAC/ILDG format and the MILC format.

Workflow and Data analysis: Data processing for lattice QCD is carried out via analysis campaigns. An analysis campaign consists of an input dataset (e.g., an ensemble of gauge configurations) and a set of interdependent processing steps (e.g., the generation of valence quark propagators, and the resulting measurements via two- and three-point correlators) that can be expressed as a directed acyclic graph (DAG). This DAG can be considered to be the workflow specification for the analysis campaign. Given the complexity of current lattice QCD analysis campaigns, which can involve hundreds of input files and thousands of intermediate and final files, it is very desirable to more closely manage these workflow specifications, and to use them to automate many aspects of executing analysis campaigns.

We propose to define a subsystem that allows workflow to be specified in a domain-specific way and later be turned into a set of instructions that can be carried out or executed on a lattice QCD compute platform. Execution includes configuration, submission, progress tracking, and accounting of an analysis campaign. It also includes input staging and storage of results. We also propose to develop a coherent data analysis package for the toolbox.

Performance analysis: As the size and complexity of the emerging high-performance computing (HPC) systems continue to grow, it is increasingly difficult to achieve a high fraction of peak performance for lattice QCD applications. Multi-core processors, complex memory hierarchies, multiple processor nodes, and complex software stacks all contribute to this difficulty, exacerbated by the rapid scaling of systems to tens of thousands of processors. To better understand lattice QCD code performance and to exploit HPC systems, we propose a set of performance studies, which will be led by the University of North Carolina computer scientists in our collaboration. Emphasis will be on: 1) performance of new generation SciDAC codes; 2) the impact of modern architectures; and 3) novel techniques for performance study using visualization.

We will develop a profiling library for QDP routines. Similar to the PQMP library, a QMP profiling library developed in SciDAC-1, the PQDP will intercept calls to QDP functions during execution and capture the performance data for such functions. It will record total time duration and the time spent in communication for each QDP call. The goal is to reveal the communication overhead for the QDP routines and to improve the overlapping of computation and communication in these routines.

We will also extend C++ support in SvPablo and conduct performance analysis for Chroma code. We will apply SvPablo and our profiling tools to carry out detailed performance studies for Chroma on various HPC systems, as was done for the MILC code, and make cross-platform performance comparisons. Furthermore, we will compare the performance of the same physics kernels running in both MILC and Chroma based on various performance metrics, and optimize the performance of these codes using SvPablo and other performance analysis tools that are being developed at the Renaissance Computing Institute.

Multiscale algorithm collaboration with TOPS: If past history is a guide, new algorithms will in the long run be as important as faster hardware in advancing research in lattice QCD. Thus, a small but essential part of developing infrastructure for the petaflop/s era should include research into fundamentally new algorithms. Indeed, one benefit of increased computational power is that by exposing more details of the short distance physics, it expands the opportunities for the use of multi-scale methods. We have begun to explore multi-level methods for QCD in collaboration with the TOPS multi-grid algorithm team.

In the early 1990's, a number of attempts were made to introduce multi-scale algorithms to QCD [40, 41, 42], which resulted in substantial theoretical progress, but failed for the most part to produce significant advantages for actual QCD simulations. However, members of our software team have recently begun working with applied mathematicians from the TOPS ISIC on this problem, and have obtained impressive preliminary results [43] using a new class [44] of adaptive algebraic multi-grid tools. It is important to continue this work, as even modest gains in performance would have a major impact on the science. In addition, Lüscher has recently introduced a blocking method based on the Schwarz alternating procedure that also shows promise. This approach too warrants further exploration. We therefore plan to continue our work on multiscale algorithms with applied mathematicians in the TOPS ISIC. To facilitate this effort we

propose to extend the capabilities of the QCD API by developing an infrastructure for multi-scale algorithms. In particular we will begin to develop at Level 4 a set of methods for rapid prototyping of multi-level algorithms. The new objects in C++ will be built on top of the QDP++ Level 2 that allows a concise paradigm to express parallelism, domain decomposition, etc.

4.3 Uniform computing environment

The Department of Energy is funding a set of terascale computers dedicated to the study of lattice QCD. These machines are being located at BNL, FNAL and JLab. There is considerable value in providing their users with a common development and job execution environment. Just as the SciDAC QCD API aims at application code portability, the uniform computing environment aims at portability within the users' working environment. This is not only a convenience, but it offers the potential to improve overall efficiency by optimizing the mix of jobs on the different architectures at the three laboratories.

Common runtime environment

One of the outputs of the SciDAC-1 lattice QCD project was the specification of a draft Common Runtime Environment [45]. This specification covers file system naming and access, the interactive environment, the batch script environment, and the parallel execution environment. In SciDAC-2 we propose to implement the QCD Common Runtime Environment at each laboratory, and to enhance and develop tools to support this specification, with a particular emphasis on meta-facility operations.

Data management: We will select or develop tools in the following areas: (1) File staging to and from the computational resource, including tools to split a single, lattice oriented file into multiple parts, and re-assemble a split (parallel) file into a single file; (2) Local file management, including migration of files to and from tertiary storage, and pinning and unpinning files; (3) Grid file management, including uploading meta-data extracted from a lattice standard file to the meta-data catalog, and domain specific graphical and command line meta-data catalog query tools. The latter will extend to retrieval and queries on the International Lattice Data Grid (ILDG).

Computational grid: Lattice QCD jobs have domain specific features which make them adjustable onto various sized parallel machines. The QMP library allows an executable to determine the size and shape of the machine on which it is running, but current batch and grid tools do not do a good job of expressing this task flexibility. Realizing this flexibility will require these developments: (1) Develop (or extend) an XML schema to describe the job's optimal machine and range of flexibility in machine parameters; (2) Modify an existing batch system scheduler, or develop a pre-processor, to deal with the flexibility in machine size, while preserving standard batch system properties (fair share, accounting, etc.). (3) Extend this onto a grid environment, with late binding to a particular resource (as opposed to the more common immediate match/binding to the currently least busy matching resource). Throughout this sub-task, efforts will be made to exploit mature grid technologies as building blocks for the lattice meta-facility.

Monitoring and controlling large systems: Lattice QCD jobs are composed of long-running, interdependent tasks. Failure of a single processor can halt progress on all processors assigned to a given job. When hardware failures occur, an application-specific set of tasks are done, such as killing the job and restarting from a checkpoint. When done manually, this approach is expensive, slow to respond, and limits scalability. We propose to develop an automated fault monitoring and mitigation system to perform these "babysitting jobs". This "Cluster Nanny" should have the following properties: (1) It should be coupled to the application. Mitigation actions depend on the properties of the application and its overall workflow. (2) It should closely monitor performance and the status of jobs, and work together with a workflow subsystem to ensure good progress for the larger analysis campaign that is being conducted. (3) It should trigger workflow re-planning, to allow for resource optimization, as components fail. This will include interactions with real-time scheduling systems. (4) It should monitor the health (performance, utilization, state) of all processors and networks in the system. In addition, tools will be developed to define the operations of lattice QCD systems and their associated fault mitigation actions. These tools will also analyze the systems and help identify single points-of-failure and resource bottlenecks.

Software for emerging hardware: As discussed in Section 5.1, this project will include the acquistion and testing of prototype hardware supporting the large procurements to be undertaken by the DOE Lattice QCD Computing Project. These hardware prototypes will include new processors and motherboards, as well

as high performance network fabrics. Software development will be necessary for the evaluation of these prototypes. Such development will include low level drivers, instrumentation for performance profiling, and the porting of hardware specific portions of the SciDAC lattice QCD libraries.

Accounting tools: A final part of the common user environment is the users' interaction with accounting systems. In the initial years of this project, users will perceive multiple accounting systems (one per site), and will likely have site specific allocations. By the third year of the project, as portable jobs are executed on the meta-facility, it will be helpful to users to have a single meta-facility allocation and view. This will require the development of a few simple tools to extend the single site accounting tools to cover multiple sites in a fault tolerant manner. One technology option is to grid enable QBank, a companion to the Maui scheduler used at both FNAL and JLab, which presents an abstraction of accounting to the scheduler.

4.4 Software task schedule

The scheduling of software tasks is give in the Gantt chart of Fig. 2. The budget requests support for 15.7 FTE per year for the software effort. This is broken down among the three main divisions of software work as follows: 4.9 FTE for Machine Specific Software, 7.7 FTE to support Infrastructure for Application Code, and 3.1 FTE for Uniform Computing Environment. These resources are nearly doubled by the contributions of physicists and software engineers at the participating institutions with no direct SciDAC support. In Appendix A.3 we briefly describe the tasks to be undertaken, the major milestones for the first two years,

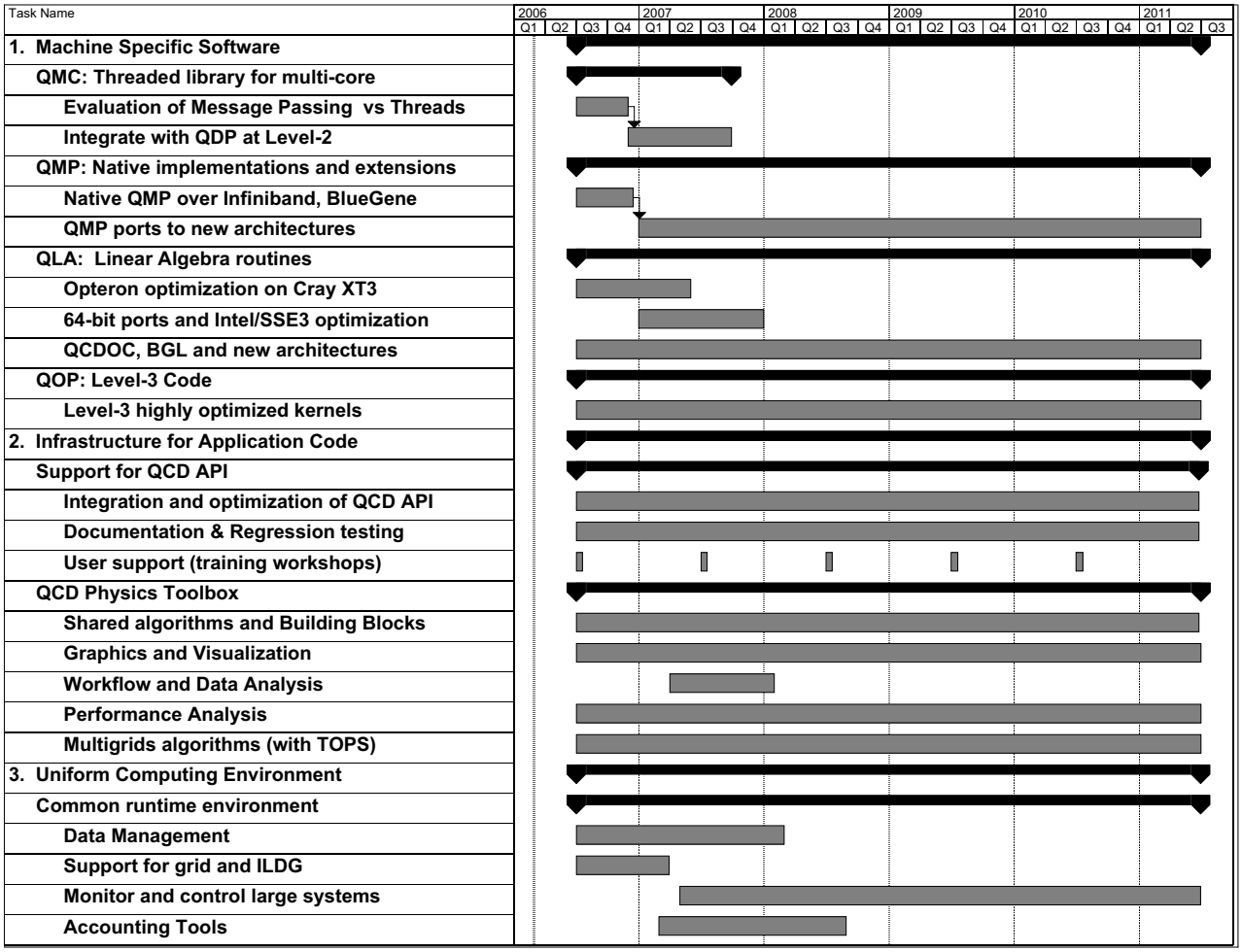

Figure 2: The schedule of software tasks.

and the FTE assignments for each participating institution.

As indicated in the Gantt chart, our software project involves both near term milestones to provide time critical software components, and long term development tasks that are to be pursued on a continuous basis throughout the grant. Important software milestones to be achieved during the first year of the grant include: (i) Design of a multi-core library interface (QMC), and the evaluation of its performance relative to simply treating each core as a separate processor. If the decision is to go forward, then the software will be developed and integration with QLA and QDP will begin; (ii) Port of QMP to Infiniband and the BlueGene/L toroidal network; (iii) Optimization of essential components of QLA for the AMD Opteron and the IBM dual core PowerPC processor; (iv) Conversion of the basic modules of the MILC code to QLA/C, introduction of templates into CPS and their restructuring in Chroma; (v) Identification of physical attributes to be visualized, cataloging of relevant data sets and design of a prototype interface; (vi) Implementation of a basic common runtime environment; and (vii) Initial design work on the workflow software and the fault monitoring and mitigation system.

The tasks that we will pursue over the full five years of the grant are equally important. They include the continuous adaptation of the low level API (QMC, QMP, QLA) to new architectures; the integration of the higher level API (QDP) into the major, freely available application codes; the definition and expansion of the common QCD physics toolbox; the design of more sophisticated algorithms with critical components optimized; and the documentation, regression testing and distribution of software libraries. As the project proceeds critical evaluations of the cost/benefit of each task may substantially alter priorities and allocation of our FTE resources. The metrics for success are a continued growth in the number of application codes written in the ACD API, the performance of these codes, and the number of physicists who use the QCD API to obtain important research results. The growth in the user community is already very encouraging.

5 Hardware Research and Development

Over the past twenty years computers specifically optimized for lattice QCD have achieved record price-performance and provided platforms for frontier physics calculations. From the Caltech Cosmic Cube to the QCDOC and the SciDAC-1 clusters, technological opportunities have been exploited to provide economical, large-scale simulation capabilities. These and similar activities carried out in Germany, Italy, Japan, the UK and the US, have also played an important role in the development of commercial supercomputers. For example, the QCDSP and QCDOC machines developed and demonstrated the architecture that is now the basis for the highly successful IBM BlueGene computers.

Work under our SciDAC-1 grant demonstrated that commodity clusters, optimized for QCD can be powerful, highly cost-effective research tools. The SciDAC-1 grant also had a large impact on the use of more specialized QCD computers, making the QCDOC machines usable by those in the U.S. lattice QCD community outside of the group of machine designers and close collaborators. We propose to continue to track the evolving commodity and semi-commodity marketplace in order to provide vital input for the parallel Lattice QCD Computing Project, and to undertake design of a fully customized successor to the QCDOC.

5.1 Investigation of Cluster Components

Background: The SciDAC-1 project undertook investigations of commodity hardware for lattice QCD. By selecting the most cost effective and appropriately balanced combinations of processor and network interconnect, as opposed to the products which individually had the best performance, and by taking advantage of the modest requirements for memory size and disk bandwidth, the SciDAC-1 project built large scale clusters dedicated to lattice QCD calculations with better price/performance than any existing general purpose parallel computing platform. The processors investigated during the project included Intel Pentium, Xeon, and Itanium, AMD Athlon and Opteron, DEC Alpha, and IBM PPC970. The project also investigated several high performance networks, including Myrinet, gigabit ethernet meshes, and Infiniband. Each year the most promising technologies were chosen to build prototype production clusters listed in Table 2.

The DOE Lattice QCD Computing Project (the “facilities project”), which started October 2005, will procure and operate large scale systems. This project is planned for FY2006 through FY2009, with funding of

Site	Cluster	Processor	Network
JLab	2m	Xeon (single)	Myrinet
JLab	3g	Xeon (single)	3D GigE
JLab	4g	Xeon (single)	5D GigE
JLab	6n	Pentium (dual core)	Infiniband
FNAL	w	Xeon (dual)	Myrinet
FNAL	qcd	Xeon (dual)	Myrinet
FNAL	pion	Pentium (single)	Infiniband

Table 2: Prototype production clusters built under SciDAC-1.

$9.2 million from the High Energy and Nuclear Physics Programs of the DOE. Approximately $6 million of this funding will be used for commodity hardware, specifically clusters in the first year, and most likely clusters for the subsequent three years. The designs of the first clusters to be built by the project in 2006 are derived directly from the prototype clusters assembled during the SciDAC-1 project. Continued hardware prototyping by the SciDAC-2 project will provide critical information for the platforms to be procured by the facilities project in FY2007–FY2009.

The prototype clusters from the SciDAC-1 project have proven to be very successful in delivering physics results. Operation for physics production of many of these clusters, specifically the 3g, 4g, and 6n clusters at JLab, and the qcd and pion clusters at FNAL, is now part of the facilities project. These clusters have an aggregate capacity of nearly two teraflop/s.

Prototyping Tasks: To support the cluster or other commodity or semi-commodity designs of the facilities project, investigations of commodity processors, chipsets, and high performance networks will be performed. These investigations will focus on those aspects most important to lattice QCD codes: memory bandwidth, floating point processing, and network performance.

During the next two years, vendor roadmaps indicate a number of technology changes which could have important benefits for lattice QCD computing. All of the principal commodity processor vendors have started the move toward multi-core designs. In 2006, dual core processors will very likely take the performance lead over single core designs, and will certainly take the lead in cost effectiveness, based on prototyping with Intel dual core processors at the end of the SciDAC-1 project. By 2007, quad core processors will be introduced.

The use of multiple processing cores will increase the floating point capabilities of commodity systems. In order to be cost effective for lattice QCD calculations, this increase in capability must be balanced by concurrent increases in memory bandwidth and improvements in network latency and bandwidth. In 2006, Intel will introduce the first systems with chipsets supporting the new fully buffered DIMM (FBDIMM) technology. Also in 2006, AMD is expected to change the integrated memory controllers on their Opteron processors to support DDR2 memories. Both of these design changes should provide significantly improved memory performance.

In addition to these mainstream processor developments, it will be valuable to track the evolution of other novel architectures, such as the Cell processor or a successor to the BlueGene/L machine. Because of the high computational capacity of these platforms, there is a possibility that either will prove to be an even more cost effective platform than clusters.

The network performance of commodity computers depends upon the input-output (I/O) bus design and on the network interfaces attached to the I/O buses. The clusters to be constructed in 2006 by the facilities project will use PCI Express (PCI-E) I/O buses and Infiniband network interfaces. Important alternatives to these buses include the higher speed PCI Express 2.0 specification, expected to appear in products in 2007, and the HyperTransport (HTX) bus, which is only available on some Opteron processor motherboards. Important alternatives to the 10 gigabit per second signal rate Infiniband fabrics used on the 2006 clusters include double and quad data rate Infiniband (16 gigabit/sec and 32 gigabit/sec data bandwidths), Infinipath, Myrinet, and Quadrics.

During each year of the SciDAC-2 project, the project will select the most important commodity or semi-commodity technologies to evaluate, that is, technologies showing the greatest promise for accelerating Lattice QCD cost effectiveness and performance. These might include new processors, new network technologies, or even novel architectures such as the Cell processor. This project will acquire modest amounts of hardware to support the porting of significant lattice QCD kernels to evaluate and optimize performance and to help determine the optimal designs for subsequent facilities project machines. The prototyping and evaluation tasks will be shared by JLab and FNAL in a complimentary way. For example, during the first year of the grant FNAL will investigate the latest AMD Opteron systems and the Pathscale Infinipath interconnect, while JLab will study the Intel dual core "Woodcrest" processor, and double data rate Infiniband fabrics. This approach leverages to the greatest extent the capabilities of these two labs, and helps to maintain the expertise needed by the facilities project. The proposed budget for hardware for each of JLab and FNAL is $40,000 per year; this is sufficient to acquire one or two small machines of sufficient size (8 or 16 nodes) to test I/O capabilities while communicating in multiple dimensions. The level of effort at each site will be 0.25 FTE.

5.2 Specialized computers for lattice QCD

Background: Over the past twenty years very substantial cost-performance benefits have resulted from the construction of computers specifically designed for lattice QCD. These performance and cost advantages have been achieved by utilizing special chips, often from the graphics market, inter-node communications strategies not available in standard commercial systems and integration/packaging targeted at the specific scale and operating environment for such QCD machines.

In order to evaluate whether special purpose machines continue to offer significant physics opportunities, we must consider the expected alternatives. The clusters planned for the LQCD Computing Project are expected to sustain several teraflop/s on production code, with scalability to tens of teraflop/s (funding constrained). On the high end, hundred-teraflop/s performance is available from BlueGene and Cray machines, and, given their success, these commercial machines should aggressively advance to the petaflop/s level. However, enormous scientific potential lies at the petaflop/s scale, and it is unlikely that in the next five years lattice QCD research budgets will approach the $100M level required to obtain dedicated computers of this scale. The lesson of the past twenty years is that innovative exploitation of trends in microelectronics, driven by the scientific imperatives of a clean, fundamental problem like lattice QCD, can offer substantial rewards at the frontier of particle and nuclear physics, and can add an important ingredient to the general advance of scientific computing.

Of course, this requires that our proposed project break new ground. Thus, we plan to begin with wide-ranging study of a number of possible directions. Presuming that a compelling approach is identified, we will then proceed to detailed design and prototype construction–activities supported in part by this SciDAC proposal. A follow-on proposal for the construction of a large-scale machine would fall outside of the SciDAC program and would be made to the base programs in High Energy and Nuclear Physics. Such a large-scale proposal would be enabled by this SciDAC-supported research and development effort, and would be driven by the scientific opportunities offered by the resulting computer.

Overall strategy: In order to justify the effort and expense associated with this design and prototyping effort, and the risks associated with the construction of a large-scale machine, substantial benefits in cost-performance must be offered by such a project. Expecting a project of this sort to require 4–5 years for completion, we must identify a direction that can yield a substantial enhancement over the expected cost-performance of commercial machines or clusters available in this time frame. Their cost performance might be optimistically predicted by applying Moore's law to the $1 per Mflop/s (sustained) in 2005 by a SciDAC-1 cluster, which over five years yields $0.1 per Mflop/s using an 18 month halving period. Thus, an appropriate target for this design effort is an order of magnitude better at that time, $0.01 per Mflop/s.

This level of cost-performance would make sustained petaflop/s available for lattice QCD in the next 4-5 years, a goal with very substantial scientific rewards. Realizing such a goal will require advances in a combination of hardware and software technology, and, very likely, a further degree of specialization in the resulting machine. Since there are scaling factors in the computational cost of generating gauge configurations which do not appear in the calculation of quantum observables on those configurations (associated

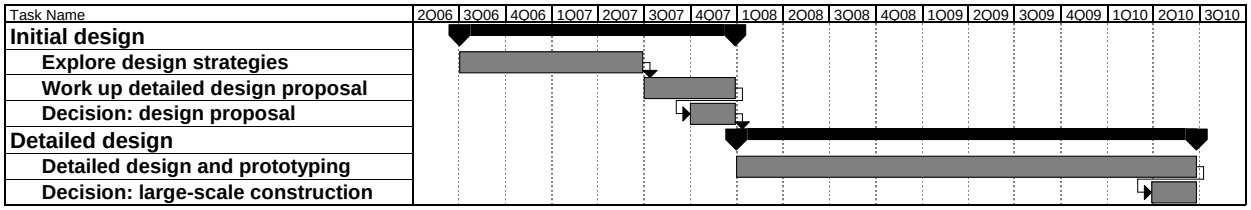

Figure 3: Proposed schedule for the design work for a next QCD machine.

with autocorrelation time and evolution step-size), the overwhelming computational needs of configuration generation will grow proportionally larger as smaller quark masses and finer lattice spacings are achieved. The generation of such gauge configurations is typically done with highly optimized code that is held stable and run for 1–2 years or more.

Thus, large scientific benefit may be realized by an innovative machine whose floating point and memory architecture yield very high performance for carefully optimized code. Even if efficient code generation for such a machine is outside the reach of current compiler technology, existing software tools (for example Peter Boyle's BAGEL code generator) demonstrate that an effective programming environment can be provided to expert users for such a machine. Thus, in order to achieve the substantial performance boost that our scientific goals demand, we should consider such architectures.

Technological opportunities: There are at least three promising directions for which we propose further study and straw man designs. The first uses a commercial "superchip" with high floating-point performance and external memory bandwidth. The SONY/IBM CELL processor, Broadcom's BCM1480 4-core chip and ClearSpeed's 50 gigaflop/s CSX600 are current examples. To be useful for QCD these chips require the design of a communications companion chip that would provide mesh communications, perhaps in 6-dimensions, and an interface to commercial memory of appropriate size and cost. The second approach exploits advances in small, low-power DSP-like cores, for example the Cortex-A8/NEON of ARM, to create a 64- to 128-processor QCD chip with substantial on-chip memory. This would be a system-on-a-chip design similar to QCDOC but more aggressive in complexity and power management. The third approach uses a very large number of small chips. These small chips could be of our own design with a single processor, multiple floating point cores, simple memory interface and support for mesh communications. Here the power and space drawbacks of such an approach may be compensated by the cost and simplicity of the chip design and the reduced requirements for the memory interface. Alternatively, this "small" chip could be a future multi-core, low-power Intel mobile chip with a specially designed companion communications chip.

As anticipated above, none of these three approaches would provide the full, integrated RISC-floating point processor environment present in the current QCDOC design. However, a price-performance point of $\approx$ \$0.01 per sustained megaflop/s may be possible with these three directions. Here we propose to develop the detailed technical designs to explore whether this is indeed possible.

Project plan: Beginning with the start of this proposed SciDAC-2 grant, we will undertake the design study outlined above. During the first year and one-half, the three technologies described above would be investigated, and one or more carried to the point of a detailed technical design whose performance, cost and risks could be reasonably established. During the final quarter of that period, this design will be reviewed by a committee appointed by the Lattice QCD Executive Committee. With input from this review, the Executive Committee will decide whether to proceed to actual design and prototyping. If a compelling proposal is made and accepted, then actual design and prototyping work would begin, culminating in a substantial prototype in two and one half years. Based on the performance of this prototype and the potential of the design to advance research in QCD, the Executive Committee will decide whether to develop a proposal to construct a large-scale machine. This design/decision flow is indicated in Gantt chart of Figure 3.

This project is of interest internationally, and we will encourage strong foreign groups with expertise and similar interests to work with us. During the first and second years, Columbia University, the lead institution for this project in the U.S., and Brookhaven National Laboratory, will coordinate these planning and design activities with the University of Edinburgh, the RIKEN BNL Research Center and Regensburg University.

From these five institutions, and possibly others which may join, we will attempt to form a design team of 5-6 principal members, and, as part of this proposal and others being made to RIKEN, PPARC and German funding agencies, we plan to add 3-4 postdoctoral-level participants.

For the first year, the costs of the proposed design effort are entirely personnel. The development of the detailed design is planned to include the procurement of development hardware, design and simulation software, as well as non-recurring engineering (NRE) costs associated with ASIC, printed circuit board and cabinet design. Funds cover a postdoctoral-level physicist dedicated to this activity at Columbia, as well as hardware, software and NRE costs. Anticipating support obtained by our possible collaboration partners in this design (RBRC and Edinburgh), costs of these final items are one-third of the total estimated on the basis of our earlier experience. These costs will become more precise as the work proceeds, and the level of support obtained at our collaborating institutions becomes known. Funding for a large-scale prototype that would be constructed at the end of this project is not included in this proposal, and will be sought outside of the SciDAC program. A final (5th year) of postdoc support is included to provide continuity between the design effort and construction of a large-scale machine.

While this effort to reach a sustained performance of $0.01/Mflop/s ($10M/sustained petaflop/s) entails considerable risk, the scientific rewards of providing this level of computational power to lattice QCD, and the influence of such a project on overall scientific computing make this a compelling direction to explore. This proposal is structured to allow the exploration to be carried out rapidly at minimal risk. The close integration of this project with the overall effort of the U.S. lattice QCD community and the software development activities described elsewhere in this proposal ensure that if this project is successful it will be of immediate and substantial benefit to the entire U.S. lattice QCD research effort.

6 Management Plan and Budget Narrative

Overall responsibility for this effort will be vested in the Lattice QCD Executive Committee, whose members (R. Brower, N. Christ, M. Creutz, P. Mackenzie, J. Negele, C. Rebbi, D. Richards, S. Sharpe, and R. Sugar) will serve as Principal Investigators. The Executive Committee sets the project's goals, and draws up plans for meeting them. It determines priorities, decides on the distribution of funds, and ensures that work is completed on schedule. At the end of each project year, it develops a rolling two year road map for specific tasks. Tasks, milestones for the first two years, and initial FTE assignments for each institution are summarized in Appendix A.3. Schedule slips of more than two months must be reported to the Executive Committee, which will then decide if a reallocation of resources or a scope change is needed. The Executive Committee has been been leading the effort to construct computational infrastructure for the U.S. lattice gauge theory community for over seven years. It holds approximately two conference calls per month, and communicates via email between calls. A consensus has been reached on nearly all issues that have come before the Executive Committee. When consensus is not reached, decisions are made by majority vote, with the Chair's vote deciding the outcome in case of a tie. The Chair of the Executive Committee, Robert Sugar, serves as spokesperson and principal contact with the Department of Energy. Each institution receiving funds under this grant has a local principal investigator, who has first level responsibility for the work carried out at his institution. The spokesperson will submit quarterly reports to the DOE on the progress of the project. He will be assisted in preparing these reports and in tracking the grant budget by Dr. Bakul Bannerjee of FNAL.

The Executive Committee has formed a number of committees to assist it in the management of the project. Their responsibilities are set out below, and a list of members of each committee is given in Appendix A.4.

Scientific Program Committee: The Scientific Program Committee monitors the scientific progress of the project, and provides leadership in setting new directions. The Committee organizes an annual meeting of the user community to review progress and obtain input on future directions. It solicits proposals for use of the dedicated computational resources available to the U.S. lattice gauge theory community: the SciDAC-1 prototype clusters, the QCDOC and the computers acquired through the LQCD Computing Project. The Committee reviews the proposals and makes preliminary allocations based on its reviews. It then organizes an open meeting of the user community to discuss the proposals and the preliminary allocations. The Committee makes final allocations following this meeting. The objective of this process is to achieve the greatest scientific benefit from the resources through broad input from the community.

Software Coordinator and Software Coordinating Committee: The Software Coordinator, Richard Brower, has overall responsibility for the software effort, providing direction and coherence to the work, and monitoring progress on all tasks. The Software Coordinator provides quarterly reports for the Executive Committee on the progress of the software effort.

The Software Coordinating Committee works with the Software Coordinator to provide overall leadership of the software effort. The Committee meets weekly in conference calls to track progress of software tasks, to discuss technical approaches for completing them, and to further clarify the tasks. The Software Coordinating Committee works in consultation with the Executive Committee to insure that critical components are available in time to keep the overall software and hardware infrastructure project on track, proposing changes in task priority and schedule to the Executive Committee as appropriate. The Software Coordinator has set up a website, http://physics.bu.edu/~brower, on which all agenda, minutes and working documents of the Software Coordinating Committee are posted, and he has also established a mail archive, (qcdapi@physics.bu.edu), for interchange of information among all members of the collaboration.

Oversight Committee: The Oversight Committee is charged with reviewing plans and priorities from the perspective of the user community, tracking progress in all aspects of the project, and making recommendations regarding alternative approaches or new directions. It meets via conference calls, which are scheduled so that the Committee can review on going progress, and provide timely advice before important decisions are taken. The Chair of the Executive Committee participates in these conference calls to obtain the advice of the Oversight Committee at first hand. The Chair of the Oversight Committee, Steven Gottlieb, maintains regular contact with all aspects of the project to keep the Committee informed of developments, and to schedule meetings appropriately.

Management of Hardware Research and Development: Donald Holmgren will oversee the investigation of cluster components, and Norman Christ the research into the design of a new specialized computer for lattice QCD. As is the case with the Software Coordinator, they will provide quarterly reports to the Executive Committee on the progress in their areas.

The overall effort will be supported by the established management structure at the three DOE laboratories (BNL, FNAL, JLab) that are major participants in the project. The project will benefit enormously from access to the SciDAC-1 clusters, the QCDOC and the computers obtained through the LQCD Computing Project, which are operated by these laboratories. This hardware is available to the entire U.S. lattice gauge theory community. The project will also benefit from the BlueGene/L computers at Boston University and MIT. All software developed under this proposal will be made publicly available, as was the software created under our SciDAC-1 grant. In addition, the large gauge configurations generated in major research projects that make use of the hardware located at BNL, FNAL and JLab will be stored in a common format, and made immediately available to the entire U.S. lattice gauge theory community in order to maximize the physics obtained from these computationally expensive data sets. This data will be be made available to the international lattice gauge theory community through the ILDG after the physicists who generate it have had an opportunity to use it in initial calculations.

Budget Narrative: The overall budget for the five year project we propose is summarized in Table 3 of Appendix A.2. The overwhelming fraction of the budget is for support of the people who will carry out the tasks discussed in this proposal. Support is requested for a total of 17.2 FTE per year of which 15.7 FTE will go into the software effort. A total of 0.5 FTE per year is requested for the investigation of cluster components, as well as $80,000 per year to purchase the components themselves. Support for 1.0 FTE per year is requested for the design of a specialized computer, and $50,000 is requested in the second year of the grant, and $125,000 in each of the third and fourth years for the procurement of hardware, design and simulation software, and non-recurring engineering costs. All funding in this project is via direct grants to participating institutions. There are no subcontracts or funded consortium arrangements.

A Appendices

A.1 References Cited

[1] The senior personnel will participate in this project or have indicated that they will make use of the infrastructure it creates are listed in Appendix A.4.

[2] The MILC Collaboration: C. Bernard *et al.*, Phys. Rev. **D64**, 054506 (2001); Phys. Rev. D **70**, 094505 (2004).

[3] The Fermilab, HPQCD, MILC, and UKQCD Collaborations: C.T.H. Davies *et al.*, Phys. Rev. Lett. **92**, 022001 (2004).

[4] The RBC Collaboration, Y. Aoki *et al.*, Phys. Rev. D**72**, 114505 (2005).

[5] The MILC Collaboration: C. Bernard *et al.*, Phys. Rev. D**70**, 114501 (2004); Nucl. Phys. (Proc. Suppl.) B140, 231 (2005).

[6] S. Gottlieb *et al.*, Proceedings of Science (Lattice 2005) 203 (2005).

[7] The HPQCD and UKQCD Collaborations: A. Gray *et al.*, Phys. Rev. **D72**, 094507 (2005).

[8] HPQCD, MILC, and UKQCD collaborations: C. Aubin, *et al.*, Phys. Rev. D **70** 031504(R) (2004).

[9] The HPQCD Collaboration: Q. Mason, *et al.*, arXiv:hep-lat/0511160.

[10] The HPQCD and UKQCD Collaborations: C. Davies *et al.*, Phys. Rev. Lett. **95**, 052002 (2005).

[11] W.J. Marciano, Phys. Rev. Lett. **93**, 231803 (2004).

[12] The Fermilab Lattice, MILC and HPQCD Collaborations: C. Aubin, *et al.*, Phys. Rev. Lett. **95** 122002 (2005)

[13] The Fermilab Lattice and MILC Collaborations: C. Aubin *et al.*, Phys. Rev. Lett. **94**, 011601 (2005).

[14] The Fermilab Lattice and UKQCD Collaborations: I.F. Allison *et al.*, Phys. Rev. Lett. **94**, 172001 (2005).

[15] F. Karsch, arXiv:hep-lat/0601013.

[16] V.G. Bornyakov *et al.*, Proceedings of Science (Lattice 2005) 157 (2005).

[17] The MILC Collaboration: C. Bernard, *et al.*, Phys. Rev. **D71**, 034504 (2005).

[18] Z. Fodor, S. D. Katz and K. K. Szabo, Phys. Lett. B **568**, 73 (2003); F. Csikor, G. I. Egri, Z. Fodor, S. D. Katz, K. K. Szabo and A. I. Toth, JHEP **0405**, 046 (2004).

[19] C. R. Allton, M. Doring, S. Ejiri, S.J. Hands, O. Kaczmarek, F. Karsch, E. Laermann, K. Redlich, Phys. Rev **D71**, 054508 (2005).

[20] P. de Forcrand and O. Philipsen, Nucl. Phys. **B642**, 290 (2002); M. D'Elia and M.-P. Lombardo, Phys. Rev. **D67**, 014505 (2003).

[21] The MILC Collaboration: C. Bernard *et al.*, Proceedings of Science (Lattice 2005) 157 (2005).

[22] C. Jung, Proceedings of Science (Lattice 2005) 150 (2005).

[23] P. Petreczky, J. Phys. Conf. Ser. **16**, 169 (2005).

[24] The LHC Collaboration: P. Hagler, *et al.*, Phys. Rev **D68**, 034505 (2003).

[25] The LHC Collaboration: P. Hagler, *et al.*, Phys. Rev. Lett. **93**, 112001 (2004).

[26] C. Alexandrou, P. de Forcrand, Th. Lippert, H. Neff, J. W. Negele, K. Schilling, W. Schroers and A. Tsapalis, Phys. Rev **D69**, 114506 (2004).

[27] C. Alexandrou, P. de Forcrand, H. Neff, J. W. Negele, W. Schroers and A. Tsapalis, Phys. Rev. Lett. **94**, 021601 (2005).

[28] The LHPC Collaboration: S. Basak, *et al.*, Phys. Rev **D72**, 094506 (2005).

[29] The LHPC Collaboration: S. Basak, *et al.*, Phys. Rev **D72**, 074501 (2005).

[30] O. Jahn, J. W. Negele and D. Sigaev, Proceedings of Science (Lattice 2005) 069 (2005).

[31] The LHPC Collaboration: R. G. Edwards, *et al.*, To be published in Phys. Rev. Lett..

[32] The LHPC Collaboration: F.D.R. Bonnet, *et al.*, Phys. Rev **D72**, 054506 (2005).

[33] The NPLQCD Collaboration: S. R. Beane, P. F. Bedaque, K. Orginos and M. J. Savage, arXiv:hep-lat/0506013.

[34] The NPLQCD Collaboration: S. R. Beane, P. F. Bedaque, K. Orginos and M. J. Savage, arXiv:hep-lat/0602010.

[35] K. Jansen, Nucl. Phys. B (Proc. Suppl.), **129**, 3 (2004); T. DeGrand, Int. J. Mod. Phys. A **19**, 1337 (2004).

[36] E. Follana, A. Hart and C.T.H. Davies (HPQCD Collaboration), Phys. Rev. Lett. **93**, 241601 (2004); S. Dürr, C. Hoelbling and U. Wenger, Phys. Rev **D70**, 094501 (2004); D.H. Adams, Phys. Rev **D72**, 114512 (2005); F. Maresca and M. Peardon, arXiv:hep-lat/0411029.

[37] Y. Shamir, Phys. Rev **D71**, 034509 (2005).

[38] The RBC Collaboration: Y. Aoki *et al.*, arXiv:hep-lat/0508011.

[39] P. A. Boyle, http://www.ph.ed.ac.uk/ paboyle/bagel/Bagel.html, 2005.

[40] R. Brower, R. Edwards, C. Rebbi and E. Vicari, Nucl. Phys. **B366**, 689 (1991).

[41] A. Hulsebos, J. Smit, and J. Vink, in Proceedings for Fermion algorithms 161, (Juelich 1991).

[42] T. Kalkreuter, J. Comput. Appl. Math. **63**, 57 (1995).

[43] J. Brannick, *et al.*, Proceedings of the 16th International Conference on Domain Decomposition Methods (2005).

[44] T. Chartier, *et al.*, SIAM J. Sci. Comput. **25**, 1 (2003).

[45] http://lqcd.jlab.org/users/RunTimeEnv.html.

A.2 Budget Summary

Table 3 below shows the total revised budget for each participating institution in each of the 4.5 years of the grant. The FY2006 budgets cover the six month period between September 15, 2006 and March 14, 2007, while the budgets for succeeding fiscal years each cover twelve month periods beginning March 15 of the fiscal year in question.

Table of Revised Budgets for SciDAC Proposal

National Computational Infrastructure for Lattice Gauge Theory

Institution	FY06	FY07	FY08	FY09	FY10
BNL	218	362	378	390	406
FNAL	249	442	456	472	485
JLab	258	458	472	484	497
Boston U.	88	183	191	198	206
DePaul U.	32	66	68	70	72
IIT	15	30	30	30	30
Indiana U.	25	51	52	54	55
MIT	113	235	244	254	264
U. Arizona	25	51	53	54	55
U. North Carolina	55	113	116	119	122
UC Santa Barbara	15	30	30	30	30
U. Utah	27	55	56	58	60
Vanderbilt U.	37	75	76	76	77
Total	1,157	2,151	2,222	2,289	2,359

Table 3: Institution and Total Budgets in $1,000

A.3 Tasks and Milestones of Participating Institutions

In this appendix we briefly describe the tasks that will be carried out by each of the collaborating institutions, the FTE budgeted for them, and indicate the major milestones for the first two years of the grant. More detailed descriptions of the tasks can be found in the work statements of the individual institutions that appear below.

BNL: BNL will continue to optimize software and implement new algorithms for the QCDOC. It will compile, install and test SciDAC software packages on this machine. BNL will continue the evolution of the Columbia Physics System (CPS) code. It will optimize the CPS for the BlueGene/L, and work on an implementation for the successor to the QCDOC. This work will continue throughout the project, although there will be greater emphasis on QCDOC software during the first three years, and on software for the QCDOC successor in the last two years. A total of 2.5 FTE is budgeted for this work.

FNAL: During the first year of the grant, FNAL will port SciDAC code from the Intel 32 bit to 64 bit environment, and will optimize the code for Opteron processors. During year one, it will also explore the approach for and determine the benefit of a native implementation of QMP over Infiniband, and if warranted, create the implementation. In collaboration with JLab and university researchers, FNAL will provide code to support multi-core processors. With computer scientists at Illinois Institute of Technology it will provide software for automated workflow, and with computer scientists at Vanderbilt it will create software to enhance the reliability of large systems. It will implement and/or deploy software to support the ILDG and other grid activities, and provide software support for the evaluation of new hardware. FNAL will work with JLab throughout the project to study commodity hardware for lattice QCD. During the first year of the grant, it will evaluate AMD Opteron processors and Pathscale Infinipath. Finally, it will assist in the overall project management. A total of 3.0 FTE per year is budgeted for these tasks.

JLab: In each year of this project JLab will carry out research aimed at improving algorithms and producing high performance code for the study of lattice QCD. During the first year of the project, JLab will focus on implementations and optimizations for multi-core processors and for the Intel/SSE3 architecture, and on support for data analysis activities. It will also expand the existing code testing framework, and provide enhanced user support in collaboration with other institutions via workshops, phone and email. JLab will work with FNAL throughout the project to study commodity hardware for lattice QCD. During the first year of the project, JLab will study the Intel dual core "Woodcrest" processor, and double data rate Infiniband fabrics. A total of 3.1 FTE per year is budgeted for these tasks.

Boston University: Boston University provides significant leadership for the project as a whole with Richard Brower serving as Software Coordinator and Claudio Rebbi as chair of the Scientific Project Committee. James Osborn of BU has special responsibility to develop the C implementation of QDP and work with collaborators at Arizona, Indiana and Utah to integrate it into the MILC code. He will also work closely with Andrew Pochinsky at MIT to optimize the QCD API for the BlueGene architecture. Brower and Rebbi are leading the physics side of the collaboration with TOPS to study multigrid methods for lattice QCD. A total of 0.97 FTE per year is budgeted for these tasks.

Columbia University: Columbia University will lead an international effort to design and prototype a specialized computer for QCD. During the first year, different design approaches will be studied, and a detailed report prepared describing the results of the study and proposing what is judged to be the best approach. During the second year, this approach will be will be pursued in greater detail, and a proposal will be submitted to the Executive Committee with a specific architecture, cost and schedule for design and construction. If this proposal is accepted, then the design and prototyping work will be pursued in subsequent years. A total of 1.0 FTE per year is budgeted for this project.

DePaul University: DePaul University will lead the design and development of a visualization tool for lattice QCD. Work will be done in collaboration with physicists involved in the project and with computer scientists at the University of North Carolina. The goals for the first year of the project are to identify and catalog the types of datasets to be visualized, identify appropriate smoothing and visualization algorithms, and develop a prototype interface. In subsequent years, plugins will be developed to read in the various types of datasets produced in lattice QCD simulations, and tools for manipulating the data in increasingly sophisticated ways will be created. A total of 1.08 FTE per year is budgeted for this effort.

University of Arizona, Indiana University and University of Utah: The MILC code is an integrated package of some 150,000 lines of scientific application code and a library of generic supporting codes, that is publicly available and widely used. Arizona, Indiana and Utah will work together to carry out a major overhaul of this code to exploit the advantages of the SciDAC software. During the first year of this effort, generic code that supports multiple science-specific applications will be converted to QLA/C to take advantage of its platform-specific optimizations. During the second year, key modules will be rewritten in QDP. Optimization and tuning of the RHMC algorithm, which is currently being incorporated into the code, will be carried out. The first production version of the algorithm will be made available by the end of year one of the grant. Production versions of the code optimized for the Cray XT3 and BlueGene/L will be completed during the first year of the grant, and multi-core and enhanced compiler improvements will be incorporated during the second year. As always, upgrades to the code will be made available to the lattice community as they are completed. Finally, improved documentation for the code will be produced and published on the web by the end of the second year of the grant. A total of 1.875 FTE per year is budgeted for this effort, divided approximately equally among the three universities.

Illinois Institute of Technology: Computer scientists at the Illinois Institute of Technology (IIT) will build a workflow management system for planning, capturing and executing LQCD analysis campaigns. This work will be done in collaboration with FNAL. During the first two years of the grant, a workflow system will be developed and integrated into the existing LQCD computing infrastructure, allowing users to describe their analysis campaign workflow through XML files or graphical interfaces, and submit them for execution. Next a scheduling system capable of interacting with the workflow system and the system performance monitor will be deployed. The final result will be an integrated workflow environment capable of handling multiple campaigns. A total of 1.083 FTE per year is budgeted for this project.

MIT: Andrew Pochinsky of MIT will lead an effort to optimize the QCD API for the BlueGene series of computers. During the first two years, the effort will focus on the BlueGene/L. The gcc compiler will be modified to make efficient use of the two arithmetic units on each processor. The QLA routines will be compiled with this modified compiler, and key routines will be hand optimized as required. A level-3 inverter for domain wall fermions will be written, and in collaboration with James Osborn of Boston University, an optimized version of QMP will be developed. This work will be aided by contract commitments made by IBM as part of the MIT purchase of a BlueGene/L. It is anticipated that in subsequent years these software developments will be extended to later models in the BlueGene line. A total of 0.925 FTE is budgeted for this effort.

University of North Carolina: Computer scientists at the University of North Carolina will develop a performance profiling library (PQDP) to analyze the performance of the MILC and Chroma codes during the first year of the project. During the second year, the PQDP will be validated by profiling the MILC code on a variety of HPC platforms, including the QCDOC, clusters, the BlueGene/L and the Cray XT3. In subsequent years the UNC SvPablo performance analysis toolkit will be extended to support analysis of C++ codes so that the PQDP can be used to study Chroma. Performance analysis will be carried out on both codes on a wide variety of HPC platforms, and a web-based performance database will be established. The goal is to optimize the performance of MILC and Chroma based on the collected performance data. Finally, UNC will work with computer scientists at DePaul on the visualization effort discussed above. A total of 0.5325 FTE per year has been budgeted for these tasks.

UCSB: As chair of the Lattice QCD Executive Committee Robert Sugar provides overall leadership and coordination of the project. UCSB will administer funds for travel not covered by grants to other participating institutions. These trips will include visits of collaboration members to participating institutions for joint work, and attendance at meetings directly related to the project. UCSB will also administer travel funds for Principal Investigators S. Sharpe and R. Sugar.

Vanderbilt University: Computer scientists at Vanderbilt will develop an automated fault monitoring and mitigation system for the large lattice QCD clusters being built at FNAL and JLab. This work will be done in collaboration with FNAL. During the first year, an integrated monitoring and control system will be designed using existing standards and tools. Also during the first year, a tool will be developed for definition of workflows, monitoring and mitigation actions, based on Vanderbilt's Generic Modeling Environment. This task will be closely coordinated with work at IIT. During the second year, model based generators will be developed to transform the designs into components and configurations for the runtime system. In subsequent years, refined versions of these tools will be developed. A total for 1.083 FTE per year is budgeted for this work.

A.4 Committees and Senior Personnel

In this appendix we list the membership of the committees making up the management team of this project. We also list the senior personnel who will participate in this project, or have indicated that they will make use of the infrastructure it creates. They comprise nearly all of the senior lattice gauge theorists in the United States, as well as computer scientists and engineers who have agreed to participate in the project.

Lattice QCD Executive Committee

Richard Brower	Boston University
Norman Christ	Columbia University
Michael Creutz	Brookhaven National Laboratory
Paul Mackenzie	Fermi National Accelerator Laboratory
John Negele	Massachusetts Institute of Technology
Claudio Rebbi	Boston University
David Richards	Thomas Jefferson National Accelerator Facility
Stephen Sharpe	University of Washington
Robert Sugar (Chair)	University of California, Santa Barbara

Scientific Program Committee

Andreas Kronfeld	Fermi National Accelerator Laboratory
Robert Mawhinney	Columbia University
Colin Morningstar	Carnegie Mellon University
John Negele	Massachusetts Institute of Technology
Claudio Rebbi (Chair)	Boston University
Stephen Sharpe	University of Washington
Doug Toussaint	University of Arizona
Frank Wilczek	Massachusetts Institute of Technology

Software Committee

Richard Brower (Chair)	Boston University
Carleton DeTar	University of Utah
Robert Edwards	Thomas Jefferson National Accelerator Facility
Donald Holmgren	Fermi National Accelerator Laboratory
Robert Mawhinney	Columbia University
Chip Watson	Thomas Jefferson National Accelerator Facility
Ying Zhang	University of North Carolina

Oversight Committee

Tanmoy Bhattacharya	Los Alamos National Laboratory
Steven Gottlieb (Chair)	Indiana University
Anna Hasenfratz	University of Colorado
Julius Kuti	University of Californa, San Diego
Robert Pennington	National Center for Supercomputer Applications
Ralph Roskies	Pittsburgh Supercomputer Center
Terry Schalk	University of California, Santa Cruz

Senior Personnel

Theodore Bapty	Vanderbilt University
Silas Beane	University of New Hampshire
Paulo Bedaque	University of Maryland
Claude Bernard	Washington University
Tanmoy Bhattacharya	Los Alamos National Laboratory
Alan Blatecky	University of North Carolina
Thomas Blum	University of Connecticut
Richard Brower	Boston University
Matthias Burkardt	New Mexico State University
Simon Catterall	Syracuse University
Shailesh Chandrasekharan	Duke University
Jie Chen	Thomas Jefferson National Accelerator Facility
Ying Chen	Thomas Jefferson National Accelerator Facility
Norman Christ	Columbia University
Joseph Christensen	McMurray University
Michael Creutz	Brookhaven National Laboratory
Christopher Dawson	Brookhaven National Laboratory
Massimo DiPierro	DePaul University
Thomas DeGrand	University of Colorado
Carleton DeTar	University of Utah
Shao-Jing Dong	University of Kentucky
Zhihua Dong	Columbia University
Terrence Draper	University of Kentucky
Patrick Dreher	Massachusetts Institute of Technology
Anthony Duncan	University of Pittsburgh
Robert Edwards	Thomas Jefferson National Accelerator Facility
E, Efstathiadis	Brookhaven National Laboratory
Estia Eichten	Fermi National Accelerator Laboratory
Michael Engelhardt	New Mexico State University
George Fleming	Yale University
Balint Joo	Thomas Jefferson National Accelerator Facility
Chulwoo Jung	Brookhaven National Laboratory
Aida El-Khadra	University of Illinois, Urbana
Rudolf Fiebig	Florida International University
Steven Gottlieb	Indiana University
Rajan Gupta	Los Alamos National Laboratory
Anna Hasenfratz	University of Colorado
Urs Heller	Florida State University
James Hetrick	University of Pacific
Ivan Horvath	University of Kentucky
Donald Holmgren	Fermi National Accelerator Laboratory
Xiangdong Ji	University of Maryland
Frithjof Karsch	Brookhaven National Laboratory
Gregory Kilcup	Ohio State University
Joseph Kiskis	University of California, Davis
Julius Kuti	University of California, San Diego
Andreas Kronfeld	Fermi National Accelerator Laboratory
Frank Lee	George Washington University
Peter Lepage	Cornell University

Keh-Fei Liu	University of Kentucky
Paul Mackenzie	Fermi National Accelerator Laboratory
Robert Mawhinney	Columbia University
Colin Morningstar	Carnegie Mellon University
Rajamani Nayayanan	Florida International University
John Negele	Massachusetts Institute of Technology
Shigemi Ohta	KEK and Riken BNL Research Center
Kostas Orginos	William & Mary University
James Osborn	Boston University
Robert Pennington	National Center for Supercomputer Applications
Peter Petreczky	Brookhaven National Laboratory
Andrew Pochinsky	Massachusetts Institute of Technology
Michael Ramsey-Muslof	California Institute of Technology
Claudio Rebbi	Boston University
Daniel Reed	University of North Carolina
Dru Renner	University of Arizona
David Richards	Thomas Jefferson National Accelerator Facility
Martin Savage	University of Washington
Stephen Sharpe	University of Washington
Junko Shigemitsu	Ohio State University
James Simone	Fermi National Accelerator Laboratory
Donald Sinclair	Argonne National Laboratory
Amarjit Soni	Brookhaven National Laboratory
Robert Sugar	University of California, Santa Barbara
Xien-He Sun	Illinois Institute of Technology
Eric Swanson	University of Pittsburgh
Chung-I Tan	Brown University
Harry Thacker	University of Virginia
Anthony W Thomas	Thomas Jefferson National Accelerator Facility
Doug Toussaint	University of Arizona
Ruth Van de Water	Fermi National Accelerator Laboratory
Steven Wallace	University of Maryland
William Watson, III	Thomas Jefferson National Accelerator Facility
Walter Wilcox	Baylor University
Ying Zhang	University of North Carolina